



Data Review and Exclusion Guidance 2.0

Revised January 2024

Prepared by:
Missouri Department of Natural Resources
Division of Environmental Quality
Water Protection Program, Watershed Protection Section

Data Review and Exclusion Guidance for Determining the Need for Permit Limits and Deriving Effluent Limits

This document provides guidance to Missouri Department of Natural Resources' Clean Water Operating Permit Section, to assist in determining when it is admissible and appropriate to exclude data before determining reasonable potential or deriving effluent limits. The establishment of these procedures for the Operating Permit Section does not preclude the department from implementing these procedures in other areas of the department, or from using alternative approaches to determine reasonable potential or to calculate effluent limits on a site-specific basis.

Determining Reasonable Potential and Deriving Water Quality-Based Effluent Limits

The purpose of the federal National Pollutant Discharge Elimination System is to address water pollution by regulating point sources that discharge pollutants to waters of the United States. As part of the permitting process, permit writers determine if specific parameters have reasonable potential to cause or contribute to an excursion above Missouri's Water Quality Standards (WQS), per 40 CFR 122.44(d)(1)(i). If reasonable potential exists, permit writers must establish a water quality-based effluent limit (WQBEL) for the applicable parameter, per 40 CFR 122.44(d)(1)(iii).

When determining reasonable potential, permit writers should use five (5) years of a facility's Discharge Monitoring Reports (DMRs) data and data values from a facility's permit application. Because data utilized for reasonable potential determination and effluent limit derivation has an important impact on water quality decisions, the data must be non-erroneous and representative of normal operations. Normal operations can include, but are not limited to, treatment adjustments, periods of optimization, and other common adjustments made by operators to address increased treatment plant loading and other common operating activities. Additionally, the data used must be representative of current facility technology. For example, if a facility upgraded from a lagoon to a mechanical plant, only the data collected following the upgrade should be used to make reasonable potential determinations.

Determining Effluent Limits and Data Exclusion for Non-Water Quality-Based Effluent Limits in the Absence of Reasonable Potential

Permit writers must also consider non-WQBELs established in regulation such as, but not limited to, 10 CSR 20-7.015 Effluent Regulations. For these effluent limits, a reasonable potential analysis cannot be determined, as reasonable potential applies only to WQBELs as defined in the Technical Support Document for Water Quality-Based Toxics Control (EPA-505/2-90/001, March 1991). In the absence of a reasonable potential analysis, the permit writer must use their best professional judgement to determine the need for other permitted effluent limits.

When conducting a determination for non-WQBELs, permit writers should review applicable rule language, the intent of the rule, facility DMRs, and other submitted data. The permit writer should review rule requirements before a determination is conducted or an effluent limit is established. Similar to the determination of reasonable potential for WQBELs, data used in the establishment of non-WQBELs must be non-erroneous and representative of normal operations. Normal operations can include, but are not limited to, treatment adjustments, periods of optimization, and other common adjustments made by operators to address increased treatment plant loading and other common operating activities. Additionally, the data used must be representative of current facility technology. For example, if a facility upgraded from a lagoon to a mechanical plant, only the data collected following the upgrade should be used to make reasonable potential determinations. Furthermore, the same determination approach should be used by permit writers in the establishment of target reduction levels. This includes, but is not limited to, the calculation of a facility's initial total phosphorus load and the calculation of a facility's overall reduction of total phosphorus utilizing representative monitoring, influent, and/or effluent data as indicated in 10 CSR 20-7.031(9)(B)2.A.

Justifiable Exclusions Based on Data Error or Non-representative Conditions

The justifiable exclusions listed below describe when data exclusion is allowable, regardless if the data are statistical outliers. These are categories to help permit writers identify circumstances where data removal is warranted; however, they are not final determinations for specific cases. The general rule is if data are erroneous or do not represent normal operations, the data are subject to removal as long as supporting information provides justification. For the data exclusion, permit writers will treat values within three standard deviations as normal operations.

Each justifiable exclusion determination results from unique circumstances; therefore, the department will consider all determinations on a case-by-case basis. Permit writers are to assist the permittee in determining if data are subject to exclusion or not. It is primarily the permittee's responsibility to notify the permit writer that data may be subject to one or more of the below justifiable exclusions, and to provide supporting data. However, if a permit writer notices a questionable data point, the permit writer may contact and work with the permittee to determine if the data point is subject to exclusion.

The justifiable exclusions are as follows:

- Quality Assurance/Quality Control (QA/QC) Exclusion: Data resulting from incorrect sampling, preservation, handling, testing, or analysis of the sample.
- Data Entry Error Exclusion: Data recorded incorrectly in the department's DMR system. It is important to replace the incorrect data with real data, if available.
- Industrial Contributor Exclusion: Data from an industrial source that is no longer discharging to the treatment facility and contributing the pollutant of concern. The permit writer will evaluate the data to determine data exclusion. Things to consider during the determination include, but are not limited to, residual strength, period of time since the facility received the discharge, and additional industrial sources with the same pollutant.
- Changes in Technology Exclusion: If the treatment system was subject to a change (e.g., upgrade or major modification) that affects the treatment of a parameter or parameters, the

permit writer can exclude the data prior to the treatment change. However, if the change in treatment does not address a given parameter, then the permit writer should not exclude the data for that parameter based solely on the change in treatment. In the event that the treatment system's actual flow increased as part of the change, the permit writer can review the parameter data to determine if treatment or dilution is causing a noticeable difference in the parameter(s) measured. Permittees should notify the permit writer of the dates of major changes so previous unrepresentative data may be excluded; the notification requirement is in accordance with 40 CFR 122.21(j), or 40 CFR 122.21(g) for industrial facilities.

- Force Majeure Exclusion: Flood, earthquake, and other natural disasters beyond the control of the technology or the operator.
- Equipment Failure Exclusion: Treatment/operational equipment failure resulting in above or below normal operations reflected in the data. Determinations for exclusion will be dependent upon the permittee meeting conditions within 40 CFR 122.41(e).
- Start-Up Exclusion: When a new treatment system begins operation or when an existing treatment system undergoes significant modifications (e.g., lagoon to mechanical plant), the DMR data for a period of time after the start-up may not depict normal operations of the treatment facility.
- Vandalism Exclusions: Data results affected by vandalism outside of the permittee's control.
- Slug Loading Exclusion: The permit writer may exclude data resulting from a slug load depending on the permittee's prevention practices.

Outliers

An outlier is a data point or points that lies outside of an expected statistical range of values in the dataset. Ignoring an outlier can skew the data or cause a problem to go unnoticed. Sometimes outliers show a significant error such as a sampling error. In other circumstances, the outlier can be real data, that is, a non-erroneous and representative data point, so the decision to exclude this data point is not a simple matter. The determination of an outlier only marks the data point as a candidate for exclusion rather than an automatic exclusion. Standard deviations and other statistical tests may be used to identify outlier data.

Standard Deviations

Standard deviation is a statistic that measures the dispersion, or spread, of data relative to its mean. Data with lower standard deviations are closer to the mean and data with higher standard deviations are farther from the mean. Standard deviation can also help in determining if data is representative of normal operations.

The empirical rule, also known as the three (3) sigma (σ) rule, states nearly all the data (99.7%) will fall within three standard deviations of the mean in a normal distribution. Using the 3σ rule, the department is reasonably certain a significant majority of data will fall within three standard deviations. Therefore, the department considers any value at or above three standard deviations of the mean as not representative of normal operations.

Additionally, the distribution of discharge data is lognormal, as data cannot be negative. In order for a permittee to support that their data is subject to the 3σ rule, the permittee will need to show

they have accounted for their data set being lognormally distributed by translating the data to a normal distribution using logarithmic calculations. After the translation, determinations regarding data meeting the 3σ rule criteria will occur.

Statistical Tests

Below are examples of statistical tests a permittee can utilize to determine outliers; however, this list is not exhaustive. If the permittee or consultant on their behalf submits an alternative outlier statistical test, the department will review the test to determine appropriateness.

Normal Distribution Data	
Grubb's Test	Tukey's Method
Generalized Extreme Studentized Deviate Test or ESD Test	Interquartile Rule for Outliers or IQR Test
MADe Method	Median Rule
2 Standard Deviation (Mean \pm 2SD)	Adjusted Boxplot
Z-Score	Tietjen-Moore test
Modified Z-Score	

Data Exclusion

The removal of data can occur prior to a reasonable potential analysis or non-WQBEL determination. The removal of a data point at this stage is not dependent on it being designated as an outlier, rather it can only be removed if it meets one of the justifiable exclusions above along with supporting information. Data removed in this way will not be utilized for non-WQBEL derivation.

Data removal may also occur after the permit writer conducts a reasonable potential analysis or non-WQBEL determination. If an outlier meets the 3σ rule criteria, it will be removed from the dataset ensuring its value does not impact effluent variability (i.e., Coefficient of Variation or CV). This approach ensures effluent limits are more appropriately calculated based on normal operations of the facility. Following the reasonable potential analysis or non-WQBEL determination, the permit writer will remove data meeting the 3σ rule criteria and calculate the CV for the remaining data, which will be used for effluent derivation.

Data exclusion either before or after the reasonable potential analysis still ensures the department is appropriately implementing 40 CFR 122.44(d)(1) by determining reasonable potential in accordance with 40 CFR 122.44(d)(1)(i), and taking into account variability along with the other criteria listed in 40 CFR 122.44(d)(1)(ii). If the parameter is found to have reasonable potential, permit writers will establish limits in accordance with 40 CFR 122.44(d)(1)(iii).

For non-WQBELs, the exclusion of data before or after the limit determination still ensures the department is appropriately implementing effluent limits in accordance with applicable rule language and the intent of the rule.

Documentation in the Factsheet and Supporting Information

If the permittee requests an exclusion, permit writers must explain in the permit's factsheet and response correspondence the decision to exclude or not exclude data. Permit writers may also review previous data, if needed, to determine if exclusion is appropriate.

The permittee must provide supporting documentation justifying any request for exclusion of data and state which category of justifiable exclusion applies. For example, if data is erroneous because of sample contamination, the permittee must provide supporting data that includes items such as duplicate test results, field blanks, lab replicates, etc., to support the contaminated sample claim. Additionally, if the permittee indicates that the outlier is due to vandalism, then providing a police report would meet the supporting information requirement.